Title of the Invention

METHOD AND APPARATUS FOR DETERMINING FORM SHEET TYPE

Inventors

Atsuhiro IMAIZUMI
Masato TERAMOTO
Tsukasa YASUE

BACKGROUND OF THE INVENTION

The present invention relates to a form sheet type determining apparatus used in a cash automatic transaction apparatus or the like, and in particular,

5 to a form sheet type determining method and apparatus for determining a type of a form sheet by reading image data of a form sheet and extracting character strings from the read image data.

Automatic machines such as a cash automatic

10 transaction apparatus and the like automatically process various kind of processes such as automatic payment using an automatic payment utilization application blank, transfer of a public charge using an account transfer blank, or paying-in transaction using

15 an ordinary deposit paying-in blank. At this time, it is necessary for the above-mentioned automatic machines to automatically determine the type of form sheets such as as an automatic payment utilization application blank or the like inserted by the user. As the

20 determining method of the form sheet type, the most general method is a method in which identifiable information such as an ID number, bar code information, a mark, etc. indicating the type of form sheet is attached to a location common to each form sheet, and

25 the form sheet type is determined by reading the

information.

Furthermore, as a determining method which does not require the above-mentioned attached information, there is known a method for determining

5 the form sheet type by reading a character string or a mark located at a specific position on the form sheet, or a method for determining the form sheet type by reading a position or a shape of a ruled line on the form sheet.

10 SUMMARY OF THE INVENTION

The method for determining the form sheet type by reading the attached information such as the ID number, bar code information, mark, etc. is an effective method only when the form sheet which is the

15 object is produced anew by laying out, however, this method cannot be applied to determine the form sheet type of already existing form sheets. Furthermore, the method for determining the form sheet type by reading the character, mark, etc. located at the specific

20 position, or the method for determining the form sheet type by reading the position and shape of the ruled line on the form sheet becomes impossible to determine the form sheet type when the layout of the form sheet is changed or the shape of the mark is changed.

25 Moreover, in these methods, there is a fear that the reading of the image becomes unstable due to a printing deviation or a variation of the scanning speed.

The object of the present invention is to solve the above-mentioned problems, and to provide an automatic determining method and apparatus of a form sheet type capable of coping with a variation of the

5 physical layout of the form sheet, and further to provide a computer program product comprising a computer usable medium having a computer readable program for executing such a method.

In order to achieve the object, in a

10 determining method of a form sheet type according to one aspect of the present invention, character strings on an input form sheet are character recognized and extracted as keywords, and these keywords are checked with respect to a matching between a plurality of sets

15 of keywords registered beforehand one set for each form sheet type, thereby to determine the type of the input form sheet.

In a determining method of a form sheet type according to one embodiment of the present invention,

20 image data of an input form sheet is read, character strings are extracted from the read image data, and each of the extracted character strings is character recognized. Then, the keywords constituted by each character string which has been character recognized

25 are respectively collated or checked for matching with sets of keywords registered beforehand, each set including keywords of each type of predetermined form sheets.

4

Furthermore, in another embodiment of the present invention, image data of a form sheet is read, and at the time of extracting character strings from the read image data, keywords constituted by each

5 character string which has been character recognized are respectively collated or checked for matching with reference character string pattern data stored in a data base, and a character string which has been character recognized and coincides at least partly with

10 any of the reference character string patterns is extracted as each keyword. The reference character string pattern data is used to extract a character recognized character string which contains a character string representing a type of the form sheets. Then,

15 the extracted keywords are collated or checked for matching with keywords intended to determine a specific form sheet type, which keywords being registered in each of the files provided for respective form sheet types, thereby to determine the type of the form sheet.

20 BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a flowchart of an automatic determining method of form sheet type in one embodiment of the present invention.

Fig. 2A-2C are diagrams useful to explain the

25 contents of a form sheet type file.

Fig. 3 is a diagram useful to explain the contents of a character pattern data base.

Fig. 4 is a diagram for explaining weight values given to keywords.

Fig. 5 is a diagram showing an example of calculation of the values of probability of form sheets.

5   Fig. 6 is a diagram useful to explain a procedure of producing a new keyword by combining extracted keywords.

Fig. 7 is a diagram showing a concrete example of producing new keywords.

10   Fig. 8 is a diagram showing a structure of an automatic determining apparatus of form sheet type in another embodiment of the present invention.

DESCRIPTION OF THE EMBODIMENTS

Hereinafter, embodiments of the present
15   invention will be described with reference to Figs. 1 to 5.

Fig. 1 is a diagram explaining a processing in an automatic determining apparatus of form sheet type according to the present embodiment. First, in
20   step S1, the keywords for determining form sheet type extracted from each of predetermined form sheets, which are the object of determination of the form sheet type determining apparatus, are registered in a file provided for each of the form sheet types.

25   Fig. 2A-2C are diagrams showing the contents of form sheets which are the object of determination of the form sheet type determining apparatus, and the

6

contents of form sheet type files in which the keywords extracted from the form sheets and used for determining the form sheet type are registered. In Figs. 2A-2C, reference numerals 1 to 3 show the form sheets, and the

5 form sheet 1 is "an automatic payment blank (bank copy)", the form sheet 2 is "an ordinary deposit paying-in slip", and the form sheet 3 is a payment blank of "electric charge". Also, reference numerals 11 to 13 show form sheet type files respectively

10 corresponding to the form sheets 1 to 3, and each of the form sheet type files includes registered therein a plurality of keywords selected from the form sheets 1 to 3 so that the types of these form sheets can be decided uniquely, and includes registered therein

15 weights respectively given to the keywords according to the degree of importance thereof. The weights are, in other words, ones dependent on the keywords themselves or keyword-specific weights.

For example, the form sheet 1 represents an

20 "automatic payment utilization application blank (bank copy)", and as the keywords, "automatic payment utilization application blank", "bank copy", "account number" and the name of blank "○△ bank" are extracted, and for the respective extracted keywords, the weight

25 values "5", "1", and "3" are given, and the file containing these given weight values together with the keywords are registered as a form sheet type file 11. That is, since the keyword "automatic payment

utilization application blank" is most important in determining the form sheet type, the weight value "5" is given.

Next, in step S2, an image picture of a form sheet 1a which is the object of determination of the form sheet type is read. The image picture is desirable to be a binary-coded picture, however, it may be a multi-value-coded picture or a color picture. Also, as a photoelectric conversion means used for reading the picture, a reading means such as a camera, a CCD sensor, etc. may be used.

Next, in step S3, all the character strings in the read image picture are extracted. In extracting the character strings, the information such as a size and a shape of concatenated pixels may be utilized.

Next, in step S4, a character recognition is performed on each of the all character strings extracted from the image picture.

Next, in step S5, the keywords which will be used as keywords for determining the form sheet type are extracted by using a character string pattern data base 31 from the character strings obtained as a result of the character recognition. The form sheet type files 11 to 13 may be stored in this data base 31.

Fig. 3 shows the contents of the character string pattern data base 31. As shown in Fig. 3, character string patterns such as "*application blank", "*charge". "*bank", "*tax", "name", "confirmation

seal", "account number", "bank copy", etc. are
registered as reference character string patterns.
Each of the character strings obtained as a result of
the above-mentioned character recognition is collated

5   or checked to determine matching with the character
string patterns registered in the character string
pattern data base 31, and a character string having at
least a part thereof coincident with any of the
character string patterns is extracted as a keyword.

10   For example, when the "automatic payment utilization
application blank (bank copy)", which is the result of
the character recognition, is collated or checked for
matching with the character string pattern "*applica-
tion blank" registered in the character string pattern

15   data base 31, it is possible to extract the "automatic
payment utilization application blank" as a keyword.
In this respect, the mark * attached to the
"*application blank", etc. indicates that all the
character strings including the "application blank" as

20   a part thereof are extracted as the keywords.

Next, in matching processing step S6, with
respect to the extracted keywords, a weight value for
the character type and a weight value for the location
are attached, and the keywords attached with these

25   weight values are collated or checked for matching with
the keywords having the weight values and registered in
the form sheet type files in step S1, and the
determination of the form sheet type is carried out

9

after obtaining a probability value.

In this step S6, first, the weight values are attached to the extracted keywords. Fig. 4 is a diagram for explaining the weight values attached to

5 the keywords. With respect to the extracted keywords, the weight values according to the character type are attached. The character type of a keyword is determined by deciding whether the keyword is a printing type or a handwritten type by detecting the

10 features such as a linearity of the well-known character string and an interval of the characters, and the weighting is performed in accordance with the determined character type. In this embodiment, since it is made a rule to use only the printing type for the

15 form sheet type determination, and not to use the handwritten type, a weight value of 1 is given when it is the printing type, and a weight value of 0 is given when it is the handwritten type.

Furthermore, the weighting is performed in

20 accordance with the described location of the extracted keyword within the form sheet, In this embodiment, as shown in Fig. 4, the form sheet is divided into 10 regions at equal interval in the vertical direction, and the character strings described in the upper potion

25 of the form sheet are regarded as being character strings which characterize the form sheet more than other character strings. Thus, the uppermost region is given a weight value of 10, and following this, weight

values 9 to 1 are given depending on the described

region of the keyword. In this respect, it is a matter

of course that the weights are given to arbitrary

locations depending on the object form sheet.

5   Next, the determination of the form sheet

type is performed. In determining the form sheet type,

the above-mentioned keywords attached with the weight

values of the character type and attached with the

weight values of the location are collated with or

10 checked to see a matching with the keywords attached

with the weight values and registered in the form sheet

type files, and the determination of the form sheet

type is performed by obtaining the value of the

probability.

15   In the present embodiment, the value of the

probability of the form sheet is obtained by using the

following calculation formulas.

K = the weight according to the character

type of the extracted keyword

20   P = the weight according to the described

location of the extracted keyword

J = the weight registered in the form sheet

type file

the value of probability = K x P x J

25   In the calculation of the value of

probability of the form sheet, the value of probability

is obtained by the above-mentioned formulas as to all

the keywords to be collated, and the total of the

obtained values is regarded as the value of probability of the form sheet, and the form sheet having the highest value of probability is determined as the form sheet type of the input picture.

5   Fig. 5 shows a calculation example of the values of probability of form sheets. In Fig. 5, it is determined that the value of probability that the type of the input form sheet is the form sheet 1, the value of probability that the type of the input form sheet is

10   the form sheet 2, and the value of probability that the type of the input form sheet is the form sheet 3 are respectively 72, 9, and 12, and the value 72 of the form sheet 1 is the largest value. Thus, the form sheet type of the input picture is determined to be the

15   form sheet 1.

          In the above-mentioned embodiment, although the form sheet type determining keywords registered in the form sheet type file in step S1 are collated or checked for matching with the keywords extracted in

20   step S5, in place of or in addition to the keywords extracted in step S5, new keywords produced by combining a plurality of sets of keywords extracted in step S5 nay be used for collation or matching.

          Fig. 6 shows a procedure for forming new

25   keywords by combining keywords mutually. In Fig. 6, reference numeral 1b denotes a form sheet which is the object of determination, and 11a denotes a form sheet type file. In producing a new keyword, first, keywords

"Heisei, ○○th year", "notification for tax payment",

... , "○✕ city" and "mayor" are extracted from the

form sheet 1b which is the determination object (step

S5). Then, the extracted keywords "Heisei, ○○th year",

5 notification for tax payment", ... , "○✕ city" and

"mayor" are combined, and for example, a new keyword

"○✕ city notification for payment of tax" 60 is

produced (step S10). Then, this new keyword is

collated or checked for matching with the form sheet

10 type determining keyword registered in the form sheet

type file 11a (step S6), thereby to determine the form

sheet type of the form sheet 1b. In this respect, the

step S10 may be performed between step S5 and step S6

in Fig. 1, or may be included in step S5.

15   Fig. 7 illustrates a method of forming new

keywords. In Fig. 7, the reference numeral 71 denotes

a group of keywords each extracted in step S5. A new

keyword is formed by combining two or a plurality of

keywords from the group of keywords 71. In this case,

20 each keyword of the group of keywords 71 is combined

with another in all manners of combination to form a

new keyword, and as a result, a group of new keywords

72 are produced.

       Fig. 8 is a block diagram showing a structure

25 of a form sheet type determining apparatus in another

embodiment of the present invention.

       In Fig. 8, a picture input portion 81 reads

an image picture of a form sheet which is the

determining object of the form sheet type determining apparatus. As a photoelectric conversion means used for the picture reading, a camera, a CCD sensor, and the like may be used.

5 A character recognition unit 82 extracts character strings from the input image picture, and performs character recognition of the extracted character strings.

A keyword extraction unit 83 extracts

10 keywords useful for form sheet type determination from the character strings obtained as a result of the character recognition.

A form sheet type determining unit (collator) 85 collates for each form sheet type file, the

15 extracted keywords with each keyword registered beforehand in the form sheet type files 11 to 13 (Fig. 2) stored in a form sheet type keyword register 86, thereby to determine the type of the form sheet.

Since the operation of the form sheet type

20 determining apparatus of the present embodiment is as described in the foregoing, the detailed explanation will be omitted here.

There will be no need to mention that the present invention can be implemented as a computer

25 usable recording medium which realizes a computer readable program code means or sequences of instructions in order to execute the form sheet type determination method described in the foregoing.

As described above, according to the above-mentioned embodiments, since the form sheets are identified by automatically extracting character strings respectively located at arbitrary positions and

5   subsequently performing character recognition to determine the form sheet type, and by collating or checking for a matching with a group of keywords described in form sheet type information, at least the following advantageous effects can be obtained.

10   It is possible to determine the form sheet type without adding new information such as a bar code, an ID number, etc. to the form sheet.

It is possible to determine the form sheet type even when the form sheet layout is changed, or the

15   font of the form sheet is changed.

It is possible to determine the form sheet type even when a printing deviation is caused in the form sheet.

It is possible to easily register the feature

20   information used to determine the form sheet. Furthermore, it is possible to reduce the storage area for storing the feature information at the time of form sheet determination.

Since the character strings at arbitrary

25   positions within the form sheet are used, the degree of freedom for performing the form sheet type determination becomes high, and at the same time, it is possible to increase the types of the form sheets which

can be determined.

It is possible to provide an automatic determining apparatus of form sheet type which can be adapted to a variation of physical layout of the form

5   sheets.

In view of the teachings described above, it is apparent that the present invention can be modified and changed in various ways. Therefore, such modifications and changes belong to the present

10   invention without departing from the scope of the present invention. For example, the form sheet type keyword register 86 may be formed as a part of the data base 31.